

# Machine Learning–Derived Predictive Risk Factors for Early Failure After Lumbar Spine Surgery: A Systematic Review and Standardized Model Framework

Oluwatomisin Enitan<sup>1,3</sup> Surendra Batudeo<sup>1</sup> Ayobami Ogunsola<sup>2</sup> Stephen Yeboah<sup>4</sup>

<sup>1</sup>Western Governors University School of Computer Science and Information Technology

<sup>2</sup>Department of Orthopedic Surgery Wake Forest University School of Medicine

<sup>3</sup>Department of Orthopedic Surgery Kwame Nkrumah University of Science and Technology

<sup>4</sup>Department of Epidemiology and Biostatistics Texas A&M University

Presenting author: Oluwatomisin Enitan MD, MSc - oenita2@wgu.edu

## Introduction

Early failure after lumbar spine surgery, defined as reoperation, major complications, or poor patient-reported outcomes (PROs), affects 10–20% of patients within two years, representing substantial clinical and economic burden. Machine learning (ML) approaches offer promise for personalized risk stratification, yet no standardized predictive model or validated clinical risk score exists for routine surgical decision-making.

## Methods

A PRISMA-guided systematic review was performed across PubMed, Embase, and Scopus (inception through February 2025). Search terms included combinations of "lumbar spine surgery," "machine learning," "predictive modeling," "artificial intelligence," "risk score," "complications," "reoperation," and "patient-reported outcomes." Inclusion criteria included studies using ML algorithms to predict postoperative outcomes in adult lumbar spine surgery. Two independent reviewers extracted data on patient demographics, comorbidities, psychosocial factors, preoperative medication use, imaging findings, surgical variables, ML model architecture, and performance metrics. Study quality was assessed using the Prediction model Risk Of Bias Assessment Tool (PROBAST). Outcomes stratified included reoperation at index level within 2 years (primary), major complications (Clavien-Dindo  $\geq 3$ ) within 90 days, poor PRO (Oswestry Disability Index improvement  $<10$  points or  $<30\%$  from baseline) at 1 year, and unplanned readmission within 30 days.

## Results

A total of 1,847 studies were identified across PubMed (n=742), Embase (n=683), and Scopus (n=422). After removal of 583 duplicates, 1,264 studies underwent abstract screening (Figure 1). Of these, 1,198 studies were excluded for reasons including non-machine learning design (456), wrong population (312), review article format (234), non-English text (89), or other reasons (107). Following full-text assessment, 13 studies met inclusion criteria and were incorporated into the final synthesis, representing 54,216 lumbar spine surgery patients. Procedures included decompression alone (31%), instrumented fusion (52%), and combined procedures (17%). Median follow-up was 24 months (range 3–60 months). Across outcomes, predictive performance was consistently strong (Figure 3). The pooled AUROC for 2-year reoperation prediction was 0.84 (95% CI 0.81–0.87), demonstrating excellent discrimination. Performance for 90-day complications averaged 0.81 (95% CI 0.78–0.84), and prediction of poor 1-year PROs achieved an AUROC of 0.77 (95% CI 0.73–0.81). Calibration was strong (slope 0.92, ideal = 1.0), with a pooled Brier score of 0.08. Model interpretation using forest plots (Figure 2) revealed a consistent set of clinical predictors. The most powerful independent predictor was preoperative opioid use  $\geq 90$  days, associated with 2.89-fold increased odds of reoperation (95% CI 2.52–3.31). Other dominant predictors replicated across  $\geq 70\%$  of studies included BMI, diabetes, age  $>65$ , smoking, depression/anxiety, ODI  $>40$ , symptom duration  $>12$  months, ASA class  $\geq 3$ , and multilevel fusion.

## Discussion

This systematic review identifies nine core predictive variables consistently associated with early failure across ML outcome studies in lumbar spine surgery. Preoperative opioid use, depression/anxiety, and baseline functional impairment demonstrated strongest associations across all outcome types. A standardized 100-point risk score framework is proposed incorporating five domains: medical factors (30 points), psychosocial factors (25 points), behavioral factors (20 points), functional factors (15 points), and surgical factors (10 points), with risk stratification categories of Low (0–30), Moderate (31–60), High (61–80), and Very High (81–100). Current evidence is limited by lack of external validation, heterogeneous outcome definitions, and inadequate assessment of surgical factors. Prospective multi-center validation with temporal external is essential for clinical deployment.

## Clinical Relevance

This review establishes the foundational evidence base for developing a unified, clinically deployable ML risk prediction model for lumbar spine surgery. The proposed framework addresses a critical gap in surgical decision-making by providing surgeons and patients with objective, personalized risk estimates to guide treatment selection, optimize modifiable preoperative factors, and establish realistic outcome expectations. Implementation could reduce avoidable surgical failures, improve patient selection, and optimize healthcare resource allocation.



